09/735,809
GAU 2614

AUSTRALIA

Patent Office
Canberra

I, KAY WARD, ACTING MANAGER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 4641 for a patent by CANON KABUSHIKI KAISHA filed on 14 December 1999.

WITNESS my hand this
Twenty-first day of December 2000

KWard

KAY WARD
<u>ACTING MANAGER EXAMINATION</u>
<u>SUPPORT AND SALES</u>

1

S&F Ref: 459938

ORIGINAL

AUSTRALIA

Patents Act 1990

<u>PROVISIONAL SPECIFICATION FOR THE INVENTION ENTITLED</u>:

Method and Apparatus for Motion Estimation

Name and Address of Applicant:

Canon Kabushiki Kaisha, incorporated in Japan, of 30-2, Shimomaruko 3-chome Ohta-ku, Tokyo, 146, Japan

Names of Inventors:

Ian Richard Gibson, Kieran Gerard Larkin and Dominic Yip

This invention is best described in the following statement:

METHOD AND APPARATUS FOR MOTION ESTIMATION

Technical Field of the Invention

The present invention relates generally to image processing and, in particular, to a method and apparatus for estimating motion.

Background Art

Known arrangements for performing motion estimation of video images typically use one of a number of methods. A first method involves estimation of illumination partial derivatives to calculate the optical flow. Another method involves the local (block based) calculation of the cross-correlation between successive images and using the cross-correlation maxima to estimate the shift between images. A third method utilizes estimation of the motion blur in any one image (using a variety of methods such as blind de-convolution, maximum likelihood estimation etc) to then use as a motion vector in optical flow.

The above arrangements suffer each from the disadvantage of being time consuming and not necessarily very robust or reliable, particularly for large magnitudes of motion.

Motion estimation is a useful tool in many aspects of image processing. Motion estimation is used in video compression (for example MPEG2). Motion estimation is also useful in identifying those portions of an image that may require updating on a display. Motion estimation also permits various portions of an image may have assigned to them, for example as metadata, motion vectors that are interpretable to provide for enhanced image processing. Motion detection may also be used for object tracking and content analysis and many other computer vision applications.

One deficiency of existing motion estimation arrangements is that they operate between adjacent frames of an image sequence, or between odd and even fields of an interlaced image within such a sequence. As a consequence, known arrangements are limited to the image sampling rate either as frames or fields. However, motion may yet occur during the sampling of, or between, such frames or fields. Known arrangements also lack fine resolution of the magnitude of motion. For high quality reproduction, which includes consideration of issues such as storage efficiency, such motion is desired to be considered.

Disclosure of the Invention

It is an object of the present invention to substantially overcome, or at least ameliorate, one or more disadvantages of existing arrangements.

According to a first aspect of the invention, there is provided a method for estimating motion of a moving object, said method comprising the steps of:

capturing at least first and second blurred images of said moving object using a capture device having a predetermined field rate, the blur in said images arising from at least motion blur of said object, said images being captured within a window period, the window period being shorter than a period defined by a field rate of the capture device;

generating an error function, said error function being a function of said first blurred image and said second blurred image;

minimising said error function; and

estimating said motion of said object from said minimised error function.

According to an other aspect of the invention, there is provided a method for estimating motion of a moving object, said method comprising the steps of:

capturing at least first and second blurred images of said moving object using a capture device having a predetermined field rate, the blur in said images arising from at least motion blur of said object, said images being captured within a window period, the window period being shorter than a period defined by a field rate of the capture device;

generating an error function, said error function comprising a cross-correlation term being a cross-correlation between said first blurred image and said second blurred image;

minimising said generated error function; and

estimating said object motion from said minimised error function.

According to another aspect of the invention, there is provided an apparatus for implementing any one of the aforementioned methods.

According to another aspect of the invention there is provided a computer program product including a computer readable medium having recorded thereon a computer program for implementing any one of the methods described above.

According to another aspect of the invention there is provided a motion estimation system comprising:

a capture device for capturing at least first and second blurred images of a moving object, the blur in said images arising from at least motion blur of said object, said images being captured within a window period, the window period being shorter than a period defined by a field rate of the capture device;

means for generating an error function, said error function comprising a cross-correlation term being a cross-correlation between said first blurred image and said second blurred image;

means for minimising said generated error function; and

means for estimating said object motion from said minimised error function.

Brief Description of the Drawings

A number of preferred embodiments of the present invention will now be described with reference to the drawings, in which:

Fig. 1 shows blur functions b_1 and b_2 representing lineal (motion) blur of spatial extent $2a_1$ and $2a_2$ respectively occurring in a direction at angle θ to the x axis;

Figs. 2A and 2B illustrate the x' components of the respective blur functions b_1 and b_2 resulting from two sequential, non-overlapping, uniform exposures with equal duration;

Figs. 3A and 3B illustrate the x' components of the respective blur functions b_1 and b_2 resulting from two uniform exposures centred on the same time, but with the one exposure twice the duration of the other;

Figs. 4A and 4B illustrate the x' component of the respective blur functions b_1 and b_2 , resulting from two uniform exposures with the same start times, but with the one exposure twice the duration of the other;

Figs. 5A and 5B illustrate the x' component of the respective blur functions b_1 and $2b_2(x)+b_1(x)$, resulting from a linear combination of b_1 and b_2;

Figs. 6A and 6B illustrate the x' components of the respective blur functions b_1 and b_2 resulting from two uniform exposures centred on the same time, but with the one exposure three times the duration of the other;

Figs. 7A and 7B illustrate yet another embodiment wherein the x' components of the respective blur functions b_1 and b_2 resulting from two non-uniform exposures centred on the same time, but with the one exposure twice the duration of the other;

Fig. 8 is a schematic block diagram of a general purpose computer upon which embodiments of the present invention can be practiced;

Fig. 9 is a schematic block diagram of an apparatus for estimating blur parameters of an image;

Fig. 10 is a schematic block diagram of a structure of an active image sensor used to capture blurred images g_1 and g_2 of Fig. 9;

Fig. 11 is a timing sequence to be used with the sensor of Fig. 10 to capture the two overlapping blurred images g_1 and g_2 with the exposure pattern of Figs. 4A and 4B; and

Figs. 12A to 12E provide comparisons between image capture according to prior art arrangements and those according to embodiments of the present invention.

Detailed Description including Best Mode

The embodiments of the invention described herein utilise the special exposure/electronic shuttering properties of a novel active image sensor in conjunction with an error minimisation scheme tailored to a sequence of images related by a specific blur relationship. In general, this method uses a combination of pure (symmetric) blur combined with pure (asymmetric) shift to obtain a maximum likelihood estimate of the underlying motion causing the blur and shift.

Consider an ideal two dimensional digital image signal, represented as a function $f(x,y)$, where x and y are the horizontal and vertical image coordinates. The image signal $f(x,y)$ is defined for all integer values of x and y, and undefined for all non-integer values of x and y. The digital image signal $f(x,y)$ is typically obtained by sampling an analog image, for instance, an image on film.

In practice, this ideal image signal $f(x,y)$ is distorted by various image degradations. In this specification image blurring is considered as well as additive noise. The image blurring to be considered is a special form of blur, namely lineal blur typically caused by motion in a plane at an oblique angle to a line of sight of a camera used for capturing the image. The blur is the projection of the object's motion into the two dimensional planeof the camera or image sensor. It is assumed that motion is approximately linear over a small period. The blurred image is represented as g_1. In the preferred embodiment of the invention, two different blurred images are used, g_1 and g_2. The blurred images g_1 and g_2 have a special relationship, examples of which are described in the embodiments. Such related images g_1 and g_2 can typically be derived using the multiple readout options that are realised with the active sensor to be described.

The blurred images g_1 and g_2 can be represented as the result of a pair of blur functions b_1 and b_2 performed on the image signal $f(x,y)$, assuming the blur is uniform over the whole image. Therefore, the blurred images g_1 and g_2 are represented using the convolution ($*$) operator as:

$$g_1(x,y)=f(x,y) * b_1(x,y) + n_1(x,y) \quad (1)$$

$$g_2(x,y)=f(x,y) * b_2(x,y) + n_2(x,y) \quad (2)$$

where n_1 and n_2 are additive noise signals.

An assumption is made that the motion is linear and uniform over the image region under consideration. The blur functions b_1 and b_2 represent lineal (motion) blur of spatial extent $2a_1$ and $2a_2$ respectively occurring in a direction at angle θ to the x axis is illustrated in Fig. 1. Consequently the blur functions b_1 and b_2 can be represented as follows:

$$b_1(x,y)=\frac{1}{2a_1}rect(\frac{x'}{2a_1}).\delta(y') \quad (3)$$

$$b_2(x,y)=\frac{1}{2a_2}rect(\frac{x'-x'_0}{2a_2}).\delta(y'-y'_0) \quad (4)$$

wherein

$$rect\left(\frac{p}{2s}\right)=0 \text{ for } p>s$$

$$1 \text{ for } |p|<s \quad \text{, and} \quad (5)$$

$\delta(p)$ is the Dirac delta function (6)

and (x',y') are rotated coordinates having the following relation:

$$\left.\begin{aligned} x' &= +x\cos(\theta)+y\sin(\theta) \\ y' &= -x\sin(\theta)+y\cos(\theta) \end{aligned}\right\} \quad (7)$$

As will be apparent from Fig. 1, the rotated coordinates have one coordinate aligned with the direction of blur. The centre of the second blur function b_2 is offset by an amount (x_0,y_0) from the first blur function b_1. In essence the blur functions b_1 and b_2 convert all point intensities in the image signal $f(x,y)$ into linear smears of length directly related to the exposure time and the image velocity.

The two dimensional Fourier transform $G(u,v)$ of an arbitrary signal $g(x,y)$ is defined as:

$$G(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\exp(-2\pi i[ux+vy])dxdy \qquad (8)$$

where u and v are the spatial frequency coordinates. Consequently, it is possible to write the transforms G_1 and G_2 of the blurred images g_1 and g_2 as:

$$G_1(u,v)=F(u,v).B_1(u,v) + N_1(u,v) \qquad (9)$$

$$G_2(u,v)=F(u,v).B_2(u,v)+N_2(u,v) \qquad (10)$$

The discrete Fourier transforms B_1 and B_2 of the blur functions b_1 and b_2 have a simple (one dimensional) form when expressed in relation to rotated frequency axes, namely:

$$\left.\begin{aligned} u' &= +u\cos(\theta) + v\sin(\theta) \\ v' &= -u\sin(\theta) + v\cos(\theta) \end{aligned}\right\} \qquad (11)$$

$$B_1(u,v) = \frac{\sin(2\pi u' a_1)}{2\pi u' a_1} \qquad (12)$$

$$B_2(u,v) = \frac{\sin(2\pi u' a_2)}{2\pi u' a_2}.\exp(-2\pi i[u' x_0']) \qquad (13)$$

From Equations (9), (10), (12) and (13), the transforms G_1 and G_2 can be rewritten as:

$$G_1(u,v) = F(u,v).\frac{\sin(2\pi u' a_1)}{2\pi u' a_1} \qquad (14)$$

$$G_2(u,v) = F(u,v).\frac{\sin(2\pi u' a_2)}{2\pi u' a_2}.\exp(-2\pi i[u' x_0']) + N \qquad (15)$$

For simplification, the noise spectra term N_1 and N_2 in Equations (9) and (10) has been substituted by a single noise spectrum term N in Equation (15). From the two

blurred images g_1 and g_2, two equations and four "unknowns" (viz the function $F(u,v)$, as well as the parameters a_1, a_2, and x'_0) are derived. By constraining the parameters a_1, a_2, and x'_0, the unknowns can be reduced to two, and a unique solution thus exists for the simultaneous Equations (14) and (15).

The above deduction can be performed upon two or more images. Embodiments using two images include:

(a) sequential, non-overlapping, uniform exposures with equal duration;

(b) both uniform exposures centred on the same time but one exposure twice the duration of the other;

(c) overlapping, uniform exposures with the same start time and one duration twice the other;

(d) variations of (b) and (c) with integer exposure multiples greater than two; and

(e) variations of (b) and (c) with non-uniform exposures. In this case the exposures are preferably representable by multiple convolutions of a uniform exposure pattern.

Case (a) above, illustrated in Figs. 2A and 2B, will not be considered further because it reduces to a well known solution of estimating pure time delay (ie. pure translational shift on the image). The blur parameter is not used in this solution. This solution only utilises the elapsed time or spatial separation between the two uniform exposures with equal duration.

Figs 3A and 3B illustrate the x' components of the respective blur functions b_1 and b_2 resulting from two uniform exposures centred on the same time, but with the one exposure twice the duration of the other, thereby representing Case (b) above. This particular relationship between the two exposures results in the following parameters:

$$\left.\begin{aligned} a_1 &= a \\ a_2 &= 2a \\ x'_0 &= 0 \\ y'_0 &= 0 \end{aligned}\right\} \qquad (16)$$

Applying these parameters to Equations (14) and (15) gives the Fourier spectra:

$$G_1(u,v) = F(u,v).\frac{\sin(2\pi u'a)}{2\pi u'a} \tag{17}$$

$$G_2(u,v) = F(u,v).\frac{\sin(4\pi u'a\ )}{4\pi u'a} + N(u,v) \tag{18}$$

Using a trigonometric expansion, a relationship between G_1 and G_2 can be found as:

$$G_2(u,v) = F(u,v).\frac{2\sin(2\pi u'a\ )\cos(2\pi u'a)}{4\pi u'a} + N(u,v) \tag{19}$$

$$G_2(u,v) = G_1(u,v).\cos(2\pi u'a) + N(u,v) \tag{20}$$

The difference between blurred images g_1 and g_2 can be measured using total modulus-squared spectral "deviation" *t*, integrated over a representative region of the spatial frequency. The modulus-squared spectral "deviation" *t* can be written as a function of the blur parameter a:

$$t(a) = \int\int_{\Re} |N(u,v)|^2 dudv = \int\int_{\Re} |G_2(u,v) - G_1(u,v).\cos(2\pi u'a)|^2 dudv\,. \tag{21}$$

The blur estimation problem can now be expressed in terms of a least-squares estimation, in which the image "discrepancy" *t(a)* is minimised with respect to the blur parameter a. The stationary condition at a minimum is simply

$$\frac{\partial t}{\partial a} = 0. \tag{22}$$

However, Equation (21) can be further expanded as:

$$t(a) = \int\int_{\Re} \left\{ |G_2|^2 + |G_1|^2 \cos^2(2\pi u'a) - \cos(2\pi u'a)(G_1 G^*{}_2 + G^*{}_1 G_2) \right\} dudv\,. \tag{23}$$

$$t(a) = \iint_{\Re} \left\{ |G_2|^2 + \frac{|G_1|^2}{2}(1+\cos(4\pi u'a)) - \cos(2\pi u'a)(G_1 G^*{}_2 + G^*{}_1 G_2) \right\} du dv . \qquad (24)$$

The blurred images g_1 and g_2 are real even functions, resulting in the Fourier spectra thereof G_1 and G_2 having Hermitian symmetry. Hence, the following apply:

$$\left.\begin{array}{l} |G_2| = real, even \\ |G_1| = real, even \\ \left(G_1 G^*{}_2 + G^*{}_1 G_2\right) = real, even \end{array}\right\} \qquad (25)$$

Equation (24) can be expressed in terms of a complex exponential, because the imaginary sine components identically sum to zero when the symmetry conditions of Equation (25) apply.

$$t(a) = \iint_{\Re} \left\{ |G_2|^2 + \frac{|G_1|^2}{2} + \frac{|G_1|^2}{2}\exp(4\pi u'a)) - \exp(2\pi u'a)(G_1 G^*{}_2 + G^*{}_1 G_2) \right\} du dv . \qquad (26)$$

Applying the coordinate transforms of Equation (11) the transformation can be rewritten as:

$$u'a = ua\cos\theta + va\sin\theta \qquad (27)$$

The image "discrepancy" t in Equation (26) is a function of two variables, and therefore $t(a, \theta)$ must be minimised with respect to both parameters, a and θ. Since a and θ have a polar structure, such allows the introduction of two new parameters X and Y with a rectilinear form as follows:

$$\left.\begin{array}{l} a\cos\theta = X \\ a\sin\theta = Y \end{array}\right\} \qquad (28)$$

Applying Equation (28) to Equation (26) results in the following:

$$t(a,\theta) = \varepsilon(X,Y) \qquad (29)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[|G_2|^2 + \frac{|G_1|^2}{2}\right]dudv$$
$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{|G_1|^2}{2}\exp(-4\pi i[uX + vY])dudv \tag{30}$$
$$- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(G_1G_2^* + G_1^*G_2)\exp(-2\pi i[uX + vY])dudv$$

The first term in Equation (30) is a constant term related to the total "energy" in the two signals g_1 and g_2, which may be ignored in order to minimise the image "discrepancy" t. The second term is the (two-dimensional) auto-correlation of blurred image g_1 which has been scaled by a factor of one half, while it's coordinates have been scaled by two. The third term is the cross-correlation of the blurred images g_1 and g_2 plus the corresponding 180 degrees rotated form.

Therefore, in the preferred embodiment, the image "discrepancy" t function can be minimised with respect to X and Y by calculating the image "discrepancy" t for all values of X and Y using the Fast Fourier Transform, which is the most efficient way to calculate large kernel correlations. If only a small blur is expected, then the correlation calculations can be limited to a small zone, and take advantage of computational efficiencies available. Additional efficiencies are possible using the symmetries apparent in Equation (30).

In another embodiment, the exposures or blur functions are chosen to have the same start times, and blur function b_2 has twice the duration of blur function b_1. Figs. 4A and 4B illustrate the x' component of the respective blur functions b_1 and b_2. This particular relationship between the two exposures results in the following parameters:

$$\left.\begin{array}{l} a_1 = a \\ a_2 = 2a \\ x_0' = a \\ y_0' = 0 \end{array}\right\} \tag{31}$$

It is noted that these exposures or blur functions b_1 and b_2 can be synthesised from Case (a) discussed above. The first exposure is just b_1, whereas the first and second exposures of Figs. 2A and 2B can be averaged to synthesise b_2 as shown in Fig. 4B.

Applying the parameters of Equations (31) to Equations (14) and (15) gives the Fourier spectra:

$$G_1(u,v) = F(u,v).\frac{\sin(2\pi u'a)}{2\pi u'a} \tag{32}$$

$$G_2(u,v) = F(u,v).\frac{\sin(4\pi u'a\ )}{4\pi u'a}.\exp\left(-2\pi i u'a\right) + N(u,v) \tag{33}$$

From Equations (32) and (33), the noise term *N(u,v)* may be expressed as follows:

$$N(u,v) = G_2(u,v) - G_1(u,v).\cos(2\pi u'a).\exp(-2\pi i u'a) \tag{34}$$

Again, the modulus-squared error or "deviation" t can be written as a function of the blur parameter a:

$$t(a) = \iint_{\Re}\left|N(u,v)\right|^2 dudv = \iint_{\Re}\left|G_2(u,v) - G_1(u,v).\cos(2\pi u'a).\exp(-2\pi i u'a)\right|^2 dudv\,. \tag{35}$$

However, Equation (35) can be further expanded as:.

$$\begin{aligned} t(a) &= \iint_{\Re}\left\{\left|G_2\right|^2 + \left|G_1\right|^2\cos^2(2\pi u'a)\right\}dudv \\ &- \iint_{\Re}\left(G_2^* G_1.\exp(-2\pi i u'a) + G_1^* G_2.\exp(2\pi i u'a)\right).\cos(2\pi u'a)dudv \end{aligned}\ . \tag{36}$$

$$\begin{aligned} &= \iint_{\Re}\left\{\left|G_2\right|^2 + \frac{\left|G_1\right|^2}{2}\right\}dudv \\ &+\frac{1}{2}\iint_{\Re}\left|G_1\right|^2\cos(4\pi u'a)dudv \\ &-\frac{1}{2}\iint_{\Re}\left(G_2^* G_1 + G_1^* G_2\right)dudv \\ &-\frac{1}{2}\iint_{\Re}\left\{G_2^* G_1.\exp(-4\pi i u'a) + G_1^* G_2.\exp(4\pi i u'a)\right\}dudv \end{aligned}\ . \tag{37}$$

Applying Equation (28) to Equation (37) results in the following:

$$
\begin{aligned}
t(a,\theta) &= \varepsilon(X,Y) \\
&= \iint_{R}\left\{|G_2|^2+\frac{|G_1|^2}{2}\right\}dudv \\
&-\frac{1}{2}\iint_{R}\{G_2^*G_1+G_1^*G_2\}dudv \\
&+\frac{1}{2}\iint_{R}|G_1|^2\cos(4\pi[uX+vY])dudv \\
&-\frac{1}{2}\iint_{R}\{G_2^*G_1.\exp(-4\pi i[uX+vY])+G_1^*G_2.\exp(4\pi i[uX+vY])\}dudv
\end{aligned}
\qquad (38)
$$

Because the blurred images g_1 and g_2 are both real functions, the following properties exist:

$$
\begin{aligned}
G_n^*(u,v) &= G_n(-u,-v) \\
|G_n(-u,-v)| &= |G_n(u,v)|
\end{aligned}
\qquad (39)
$$

The full (or at least symmetrical) limits in frequency space allows the last two terms of Equation (38) to be joined.

$$
\begin{aligned}
&\varepsilon(X,Y) = \\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{|G_2(u,v)|^2+\frac{1}{2}|G_1(u,v)|^2\right\}dudv \\
&-\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{G_2^*(u,v)G_1(u,v)+G_1^*(u,v)G_2(u,v)\}dudv . \\
&+\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|G_1(u,v)|^2\exp(4\pi[uX+vY])dudv \\
&-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}G_1(-u,-v)G_2(u,v).\exp(4\pi i[uX+vY])dudv
\end{aligned}
\qquad (40)
$$

Only the last two terms of Equation (40) are functions of X and Y. These are therefore the only terms determining where the minimum of the deviation function $\varepsilon(X,Y)$ is located. In particular, the last two terms can be recognised as having the form of auto and cross-correlation functions).

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|G_1(u,v)|^2\exp(4\pi[uX+vY])dudv$$
$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}G_1(-u,-v)G_2(u,v).\exp(4\pi i[uX+vY])dudv \quad . \qquad (41)$$
$$=\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}G_1(-u,-v).\{G_1(-u,-v)-2G_2(u,v)\}.\exp(4\pi i[uX+vY])dudv$$

Equation (41) represents the convolution of $g_1(-x,-y)$ with the difference of $(g_1 -x,-y)$ and $2g_2(x,y)$ evaluated at coordinates (X',Y'), where

$$\left.\begin{aligned} X' &= 2X \\ Y' &= 2Y \end{aligned}\right\} \qquad (42)$$

Case (c) is special case where:

$$\left.\begin{aligned} f(x,y) &= \delta(x,y) \\ b_1(x,y) &= g_1(x,y) \\ b_2(x,y) &= g_2(x,y) \end{aligned}\right\} \qquad (43)$$

and is illustrated in Figs. 5A and 5B. An alternative solution is to negate the second term in Equation (41), and find the maximum instead. This special case is equivalent to finding the maximum cross-correlation for the blur functions b_1 and b_2 shown in Figs. 2A and 2B, with the absolute blur functions shifted by a. However, this equal shift of both blur functions does not effect the cross-correlation between the blur functions b_1 and b_2. Thus, this special case reduces to the conventional shift solution.

A person skilled in the art would recognise that many variations of the above embodiments exist. For example, the blur function b_2 can be chosen such that its duration comprises multiples of the duration of blur function b_1, wherein the multiples are greater than two which correspond with the embodiments described above.

By considering embodiments with higher durations for blur function b_2 relative to the blur function b_1, a sequence of auto-correlations and cross-correlations are evaluated with various integer re-scaling factors. However, the overall method is similar in that a direct calculation of the minimisation is possible using fast correlation techniques.

In order to illustrate the above, yet another embodiment is described. Figs. 6A and 6B illustrate the x' components of the respective blur functions b_1 and b_2 resulting from two uniform exposures centred on the same time, but with the one exposure three times the duration of the other. An alternative embodiment with the same start time can be analysed in a very similar fashion and will not be explicitly derived.

This particular relationship between the two exposures results in the following parameters:

$$\left.\begin{aligned} a_1 &= a \\ a_2 &= 3a \\ x_0' &= 0 \\ y_0' &= 0 \end{aligned}\right\} \qquad (44)$$

Applying the parameters of Equations (44) to Equations (12) and (13) gives the Fourier spectra:

$$G_1(u,v) = F(u,v).\frac{\sin(2\pi u'a)}{2\pi u'a} \qquad (45)$$

$$G_2(u,v) = F(u,v).\frac{\sin(6\pi u'a\ )}{6\pi u'a} + N(u,v) \qquad (46)$$

Using the following trigonometric expansion:

$$\sin(6\pi u'a) = \left[4\cos^2(2\pi u'a) - 1\right]\sin(2\pi u'a) = \left[2\cos(4\pi u'a) + 1\right]\sin(2\pi u'a). \qquad (47)$$

the following total modulus-squared spectral "deviation" t, integrated over a representative region of the spatial frequency, must be minimised:

$$t(a) = \iint_{\Re} |N(u,v)|^2\, dudv = \iint_{\Re} |G_2(u,v) - G_1(u,v).\cos(2\pi u'a)|^2\, dudv. \qquad (48)$$

Expanding Equation (48) gives:

$$t(a) =$$

$$\int\int_R \left\{ |G_2|^2 + \left[G_1G_2^* + G_2G_1^*\right]\left[2\cos(4\pi u'a)+1\right]\Big/_3 + |G_1|^2\left[2\cos(8\pi u'a) + 2\cos(4\pi u'a) + 3)\right]\Big/_9 \right\} dudv \tag{49}$$

Thus, by examining each of the terms above, the blur parameter can be directly estimated, in a manner similar to the previous embodiments. Equation (49) includes two cross-correlations re-scaled by factor two and two auto-correlations re-scaled by factors four and two respectively.

When blur functions are considered with the duration of one of the blur functions a larger integer value of that of the other, the number of re-scaled auto- and cross-correlations to be evaluated also increase, but otherwise similar relationships exist.

Non-uniform exposures can similarly be used. Figs. 7A and 7B illustrate yet another embodiment wherein the x' components of the respective blur functions b_1 and b_2 resulting from two non-uniform exposures centred on the same time, but with the one exposure twice the duration of the other. An alternative embodiment with the same start time can be analysed in a very similar fashion and will therefore not be explicitly derived.

The typical non-uniform exposure pattern is derived from just one convolution. The method is readily extended to multiple convolutions.

The Fourier spectra in this embodiment is as follows:

$$G_1(u,v) = F(u,v).\left[\frac{\sin(2\pi u'a)}{2\pi u'a}\right]^m \tag{50}$$

$$G_2(u,v) = F(u,v).\left[\frac{\sin(4\pi u'a\ )}{4\pi u'a}\right]^m + N(u,v) \tag{51}$$

wherein the integer value m represents the number of times a uniform exposure pattern has to be convolved to produce the required non-uniform exposure pattern. The embodiment illustrated in Figs. 7A and &B was derived using m=2.

Minimisation needs to be applied to the following total modulus-squared spectral "deviation" t:

$$t(a)=\iint_{\Re}\left|N(u,v)\right|^{2}dudv=\iint_{\Re}\left|G_{2}(u,v)-G_{1}(u,v).\cos^{m}(2\pi u'a)\right|^{2}dudv. \qquad (52)$$

Using similar argument to those in previous embodiments, Equation (52) will, in general, expand into a series of auto- and cross-correlation with various integer re-scaling factors:

$$\begin{aligned}t(a)=&\iint_{\Re}\left\{|G_2|^2+|G_1|^2\cos^{2m}(2\pi u'a)\right\}dudv\\&-\iint_{\Re}\left(G_2^*G_1+G_1^*G_2\right)\cos^{m}(2\pi u'a)dudv\end{aligned}. \qquad (53)$$

Using m=2 gives:

$$\begin{aligned}t(a)=&\iint_{\Re}\left\{|G_2|^2+|G_1|^2\left[\cos(8\pi u'a)+4\cos(4\pi u'a)+3\right]\Big/8\right\}dudv\\&-\iint_{\Re}\left\{\left(G_2^*G_1+G_1^*G_2\right)\left[\cos(4\pi u'a)+1\right]/2\right\}dudv\end{aligned}. \qquad (54)$$

Specifically in this the case with m=2, from Equation 54, the cross-correlation is re-scaled by a factor of two, and the auto-correlation re-scaled by factors of two and four.

The foregoing provide direct computational methods for estimating blur parameters over an infinite plane. For general implementation, the algorithms are desirably modified for use with finite and discrete data-sets typical of digitally sample images. The computation may be modified to work over regions, typically rectangular blocks of pixels, which are significantly larger than the blur parameter sought to be estimated, yet small enough so that the blur parameter remains substantially constant over the region. Also, edge effects which are not present in the infinite plane, may be mitigated by suitable image padding and windowing methods. These constraints are typical of other correlation-based image analysis techniques and consequently need not be discussed further herein.

Fig. 9 illustrates a block diagram of apparatus 9 for estimating the blur parameters of blurred images according to an embodiment of the present invention. The apparatus 9 has an active image sensor 10, preferably formed in CMOS technology, for capturing the blurred images g_1 and g_2, a plurality of correlators 20 and 30 for performing autocorrelation of the blurred image g_1 and cross-correlation between the two images g_1 and g_2 respectively, an error function calculator 40 for evaluating an error function t over

all possible displacements using the results from the correlators 20 and 30, and an extreme locater 50 for finding the displacement with the minimum value for the error function *t*.

The operation of the image sensor 10, as implemented with CMOS technology, will be further explained with references to Figs. 10 and 11. Fig. 10 shows a circuit layout of a part 12 of the image sensor 10, the part 12 being arranged to capture a single pixel of the digital image signal $f(x,y)$ by means of a pixel cell 60.

The pixel cell 60 is formed of a photodiode 63 which acts as an image pixel sensor and which connects to a transistor 64 to discharge a parasitic capacitor Cp according to the intensity of illumination falling upon the photodiode 63. The parasitic capacitor Cp may be formed by a floating diffusion. A transistor switch 62, operated by a reset signal 'res', acts to reset the capacitor Cp by applying a supply rail potential to the capacitor Cp. A signal "TX" is applied to the parasitic capacitor Cp enabling the capacitor Cp to discharge through the operation of the photodiode 63. It will be appreciated therefore that the sensor 10 operates according to negative logic principles where the charge on the capacitor Cp is the inverse of the light intensity-duration function applied to the photodiode 63. A select signal 'sel' operates a transistor switch 65a which, together with a source-follower transistor 65b, couples the voltage on the parasitic capacitor Cp to a pixel data bus 61. The duration that the signal 'TX' is applied is usually called the exposure time, and the charge of the parasitic capacitor Cp is the value of the pixel cell 60. A number of the pixel cells 60 may be connected to the pixel data bus 61.

In order to provide a sequence of output pixel values of an image, a switching arrangement 66 is provided to couple such values to an output 67 of the part 12. The sequence may, for example, be a raster sequence.

Signals "N1" and "N2" are applied to transistor switches SW1 and SW3, causing them to conduct thereby providing for noise charge values to be stored on capacitors C1 and C3 respectively. Such noise values are representative of circuit noise and are available on the pixel data bus 61 when the 'sel' signal of each of the cells 60 connected to the bus 61 are all disabled or not selected.

Signals 'S1' and 'S2' are applied to turn ON switches SW0 or SW2 respectively, so that capacitors C0 and C2 are selectively charged according to the pixel value as present on the pixel bus 61.

Signals 'HST0'and 'HST1" are applied to turn ON switch pairs SW4 and SW5, and SW6 and SW7 respectively. The switches SW5 and SW7 act to couple the noise values

on the capacitors C1 and C3 respectively to a buffer 'bufn', while the pixel values present on the capacitors C0 and C2 are coupled via operation of the switches SW4 and SW6 respectively to a buffer 'bufs'. A differential amplifier 'diffamp', connected to the buffers 'bufn' and 'bufs', finds the difference between the two signals to output a signal indicative of the light fallen upon the photodiode 63.

By forming the image sensor 10 as a matrix of pixel cells 60 formed in rows (lines) and columns, overlapped double sampling of an image that gives rise to blur functions illustrated in Figs. 4A and 4B, can be achieved using the circuit 12 of Fig. 10 by applying to the circuit 12 a signal sequence as shown in Fig. 11.

When a signal 'sel0' is asserted, all the pixel cells 60 in row 0 of the matrix are enabled, and each drive a corresponding pixel data bus 61 in the corresponding column.

The signal 'res' is then asserted, causing all the pixel cells 60 in line 0 to be reset, and leaving each cell 60 driving a noise value on the corresponding pixel data bus 61.

As the signal 'res' id de-asserted, the signals 'N1' and 'N2' are then asserted to latch the noise values into capacitors C1 and C3. The signal 'TX' is then asserted for a period 'Tex'. During the period 'Tex', all the pixel cells 60 in line 0 discharge their corresponding parasitic capacitors Cp according to the light intensity on line 0. The signal 'S1' is asserted in the middle of the exposure time to store the pixel values at that time in the respective capacitors C0. At the end of the exposure time 'Tex', the pixel values at that time are stored in capacitors C2 by asserting signal 'S2'. The signal 'HST0' is then asserted to switch on SW4 and SW5 to find the true values of the pixels during the first half of exposure time 'Tex'. The signal 'HST1' is asserted to switch on SW6 and SW7 to find the true values of the pixels during the second half of exposure time 'Tex'. The signals 'HST0' and 'HST1' are asserted progressively along line 0 for each pixel to output a sequence of pixel values for the line. This is repeated until all the pixels in line 0 are read.

Such a mode of operation results in a pixel sequence as follows, for n pixels in the line:

Pixel_1_half, Pixel_1_full, Pixel_2_half, Pixel_2_full, ... Pixel_n_full.

A next line (line 1) is selected by asserting 'sel1', and the above is repeated.

Alternatively, the mode of operation may be altered to provide the following sequence:

Pixel_1_half, Pixel-2_half ... Pixel_n_half, Pixel 1_full, Pixel 2_full, ... Pixel_n_full.

It will be apparent from the forgoing that where S1 is activated at the middle of TX and S2 at the end, the first image obtained by virtue of S1, for a stationary target which presents no blur, will have half the intensity of the second image obtained at S2. As such, any image derived at S1 will include, for all intents and purposes, the same components of the image obtained at S2. This is to be contrasted with prior art arrangements where single frames or fields are captured one after the other in which case the entire sensor reset and newly illuminated.

Once the blur parameters have been estimated from the two sub-frame images, a new, de-blurred image may be generated. This will not be required for typical video image sequences, because the sub-frame blur will be much smaller than the inter-frame movement blur. However, if a still image with minimal motion blur is required, then various de-convolution methods may be used to generate such an image. Such methods can utilise the estimated blur parameters to obtain a direct solution. Alternatively, iterative methods may be used with a first iteration using the estimated blur parameters to accelerate convergence.

By utilizing sub-frame or sub-field sampling of an image through the arrangement of Fig. 10, it is thus possible to obtain a motion vector from the equivalent of single prior art image sample, whereas according to prior art configurations, at least two traditional samples would be required. Thus the described embodiment provides for an estimation of motion blur for each traditional image capture, be it a frame or an interlaced field.

The above can be readily seen from Figs. 12A to 12E. Fig. 12A shows a prior art video signal 200 where an image frame is captured during period 202 and output from a prior art sensor during a period 204 prior to the next frame 202 being captured. Fig. 10B shows a similar signal 206 for an interlaced frame where a first field 208 is captured and output over a period 210. A second field 212 is captured and output over a period 214 prior to the first field of the next frame being captured.

According to an embodiment of the invention, as shown in Fig. 12C, a first image 222 is captured immediately followed by a second image 224, the second image 224 for example corresponding to that captured during the first field 208 of Fig. 12B. The data of the two images 222 and 226 is output during a period 226 before images 228 and 230 are captured for the next field and output during the period 220. It follows from the above, that the field period of the capture device will be defined at least by a time taken to read out information of twice number of total pixels of an image sensor of the capture device, so that for each field period, at least one estimate of motion is made.

Whilst Fig. 12C shows the motion blur images being captured one immediately after the other, such need not always be the case. As seen in Fig. 12D, a capture sequence 234 includes the capture of a first blur image 236 overlapping with the capture of a second blur image 238. In Fig. 12E, a capture sequence 240 involves capturing one image 242 over a period twice that of another image 244.

The blur image processing methods described above and schematically illustrated in Fig. 9 may be practiced using a conventional general-purpose computer system 100, such as that shown in Fig. 8 wherein the Fourier Transform as well as the auto- correlation and cross-correlation calculations, error function calculation and extreme location may be implemented as software, such as an application program executing within the computer system 100. In particular, those calculations are effected by instructions in the software that are carried out by the computer. The software may be stored in a computer readable medium, including the storage devices described below. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable medium having such software or computer program recorded on it is a computer program product.

The computer system 100 comprises a computer module 102, input devices such as a keyboard 110 and mouse 112, output devices including a display device 104. The computer module 102 typically includes at least one processor unit 114, a memory unit 118, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 122, and an I/O interface 116 for the keyboard 110 and mouse 112 and peripherals interface 132, to which an appropriate image sensor 134 configured according to Figs. 10 and 11 is connected to provide image pixel data to the computer module 102. A storage device 124 is provided and typically includes a hard disk drive 126 and a floppy disk drive 128. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 120 is typically provided as a non-volatile source of data. The components 114 to 128 of the computer module 102, typically communicate via an interconnected bus 130 and in a manner which results in a conventional mode of operation of the computer system 100 known to those in the relevant art. Examples of computers on which the embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

Typically, the application program of the preferred embodiment is resident on the hard disk drive 126 and read and controlled in its execution by the processor 114.

Intermediate storage of the program may be accomplished using the semiconductor memory 118, possibly in concert with the hard disk drive 128. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 120 or 128. Still further, the software can also be loaded into the computer system 100 from other computer readable medium including magnetic tape, a ROM or integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 102 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including email transmissions and information recorded on websites and the like. The foregoing is merely exemplary of relevant computer readable mediums. Other computer readable mediums may be practiced without departing from the scope and spirit of the invention.

The calculation of the Fourier transforms as well as the auto-correlation and the cross-correlation may alternatively be implemented in dedicated hardware such as one or more integrated circuits performing these functions or parts thereof. Such dedicated hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories.

The embodiments described permit the estimation of motion of an object captured in an image through obtaining at least two blur images where traditionally only a single image would be obtained. The blur images are then processed to determine motion parameters across the pair of images. The motion parameters may then be associated with one of the images, usually the second of the two, two provide an image with corresponding motion parameters obtained during the generation of the (second) image. The motion parameters may then be used as desired for object identification within the image, for example. Where the images for a sequence of such images, the motion parameters permit the tracking or other monitoring of moving objects contained within the image sequence. Motion detection using an embodiment of the present invention is desirable because such associates one motion vector map to each frame or field, which not only is smaller in magnitude thereby affording finer resolution, but also a more accurate representation of motion.

The foregoing describes only some embodiments of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the invention, the embodiments being illustrative and not restrictive.

In the context of this specification, the word "comprising" means "including principally but not necessarily solely" or "having" or "including" and not "consisting only

of". Variations of the word comprising, such as "comprise" and "comprises" have corresponding meanings.

The claims defining the invention are as follows:

1. A method for estimating motion of a moving object, said method comprising the steps of:

capturing at least first and second blurred images of said moving object using a capture device having a predetermined field rate, the blur in said images arising from at least motion blur of said object, said images being captured within a window period, the window period being shorter than a period defined by a field rate of the capture device;

generating an error function, said error function being a function of said first blurred image and said second blurred image;

minimising said error function; and

estimating said motion of said object from said minimised error function.

2. A method according to claim 1, wherein said images are captured consecutively, one immediately after the other, within said window period.

3. A method according to claim 1, wherein said images are captured sequentially within said window period with a time difference between a start time of capture of said first blurred image and a start time of capture of said second blurred image.

4. A method according to claim 1, wherein a start time of capture of the first blurred image and a start time of capture of the second blurred image are concurrent.

5. A method according to claim 2 or 3, wherein said first blurred image has an exposure duration substantially equal to an exposure duration of the second blurred image.

6. A method according to claim 3 or 4, wherein said second blurred image has an exposure duration that is a predetermined integer multiple of an exposure duration of said first blurred image.

7. A method according to claim 3 or 4, wherein said exposure durations of said first and second blurred images overlap.

8. A method according to any one of the preceding claims, wherein said field rate of the capture device is defined at least by a time taken to read out information of half the number of total pixels of an image sensor of the capture device, so that for each corresponding field period, at least one estimate of motion is made.

9. A method according to any one of the claims 1 to 7, wherein said field rate of the capture device is a rate at which plural frames of a video sequence are captured (ie. the frame rate).

10. A method according to any one of claims 1 to 7, wherein an exposure pattern (profile) of said exposure duration of at least one of said blurred images is non-uniform.

11. A method according to claim 10, wherein said exposure pattern (profile) comprises a triangular profile.

12. A method for estimating motion of a moving object, said method comprising the steps of:

capturing at least first and second blurred images of said moving object using a capture device having a predetermined field rate, the blur in said images arising from at least motion blur of said object, said images being captured within a window period, the window period being shorter than a period defined by a field rate of the capture device;

generating an error function, said error function comprising a cross-correlation term being a cross-correlation between said first blurred image and said second blurred image;

minimising said generated error function; and

estimating said object motion from said minimised error function.

13. A method according to claim 10, wherein said error function further comprises an auto-correlation term being an auto-correlation of said first blurred image.

14. Apparatus configured to perform the method of any one of the preceding claims.

15. A motion estimation system comprising:

a capture device for capturing at least first and second blurred images of a moving object, the blur in said images arising from at least motion blur of said object, said images being captured within a window period, the window period being shorter than a period defined by a field rate of the capture device;

means for generating an error function, said error function being a function of said first blurred image and said second blurred image;

means for minimising said error function; and

means for estimating said motion of said object from said minimised error function.

16. A motion estimation system comprising:

a capture device for capturing at least first and second blurred images of a moving object, the blur in said images arising from at least motion blur of said object, said images being captured within a window period, the window period being shorter than a period defined by a field rate of the capture device;

means for generating an error function, said error function comprising a cross-correlation term being a cross-correlation between said first blurred image and said second blurred image;

means for minimising said generated error function; and

means for estimating said object motion from said minimised error function.

17. A system according to claim 15 or 16 wherein said means for generating, said means for minimising and said means for estimation collectively comprise a computer system incorporating a sequence of program instructions for estimating said motion using said images output from said capture device.

18. A method of estimating motion of an object substantially as described herein with reference to any one of the embodiments as that embodiment is illustrated in the drawings.

19. Apparatus for performing the method of claim 18.

20. A computer readable medium comprising a computer program product for performing the method of claim 18.

Dated 14 December, 1999
Canon Kabushiki Kaisha

Patent Attorneys for the Applicant/Nominated Person
SPRUSON & FERGUSON

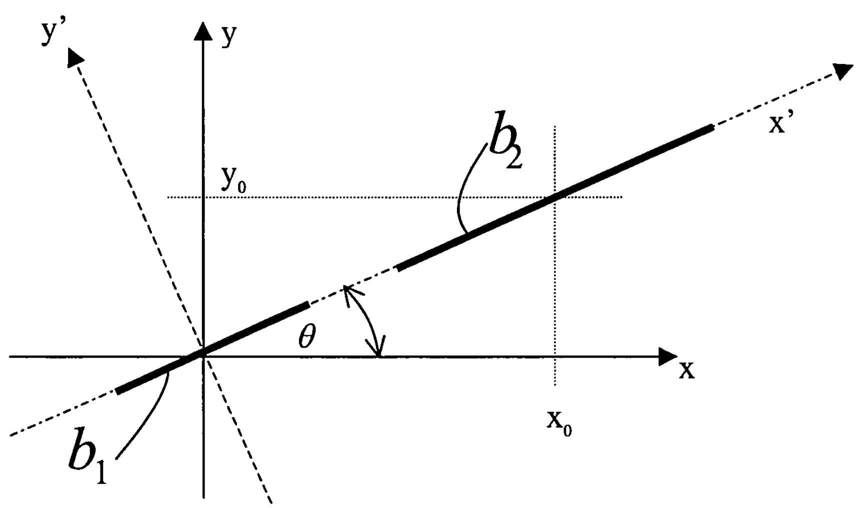


Fig. 1

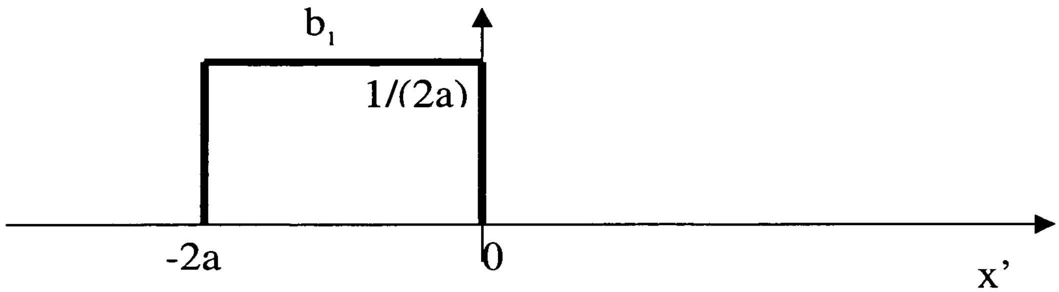


Fig. 2A

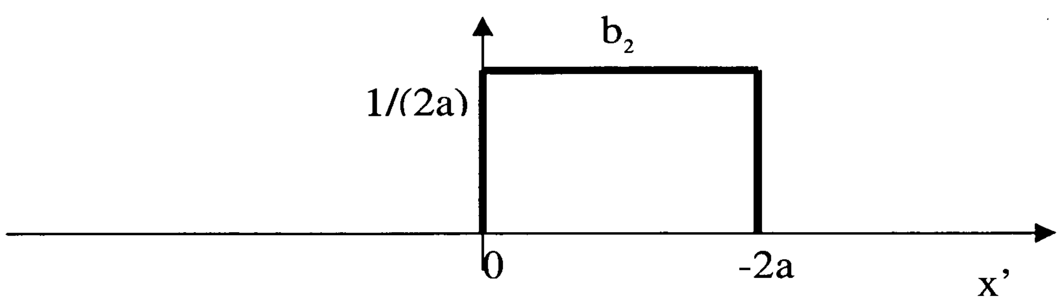


Fig. 2B

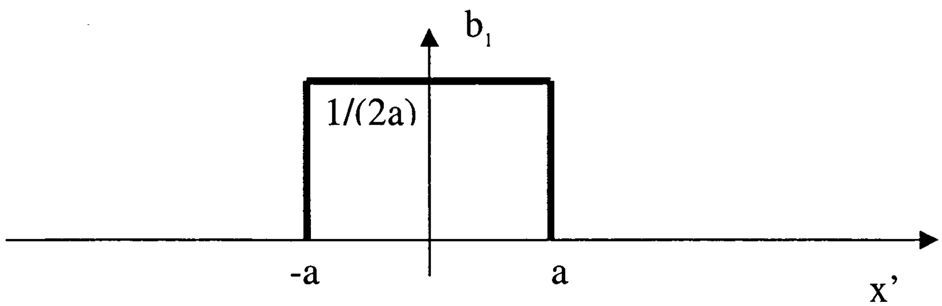


Fig. 3A

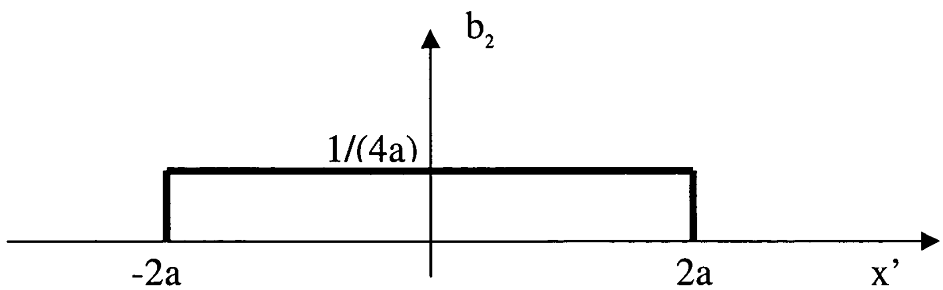


Fig. 3B

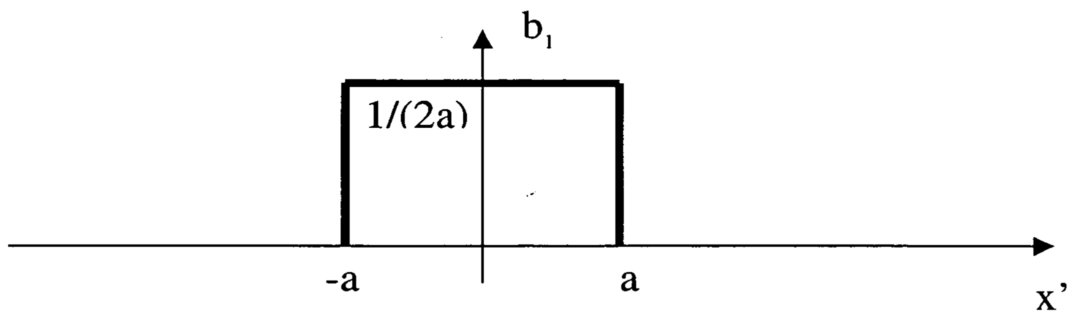


Fig 4A

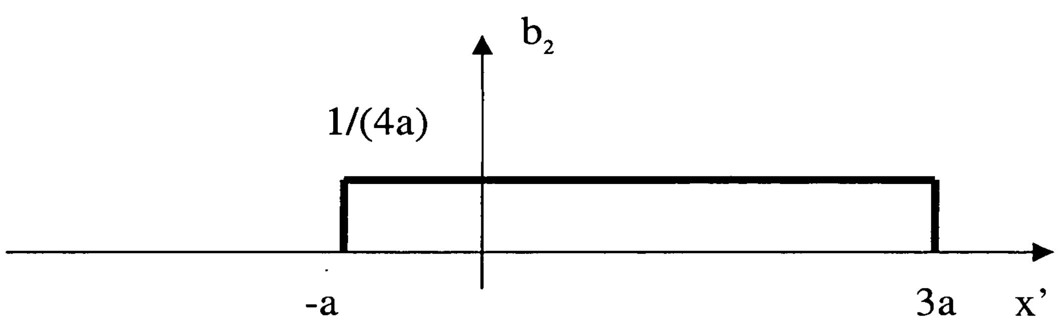


Fig 4B

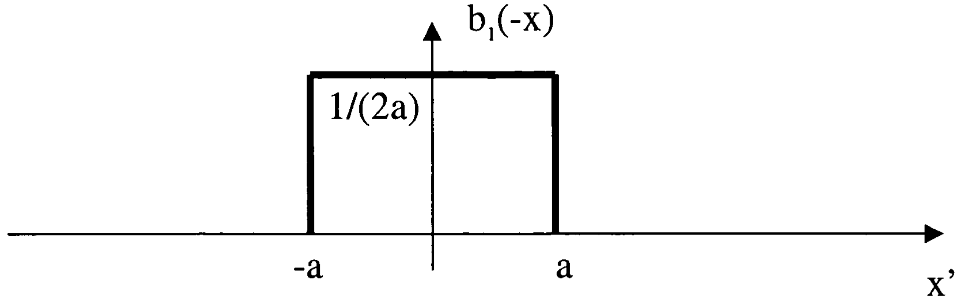


Fig 5A

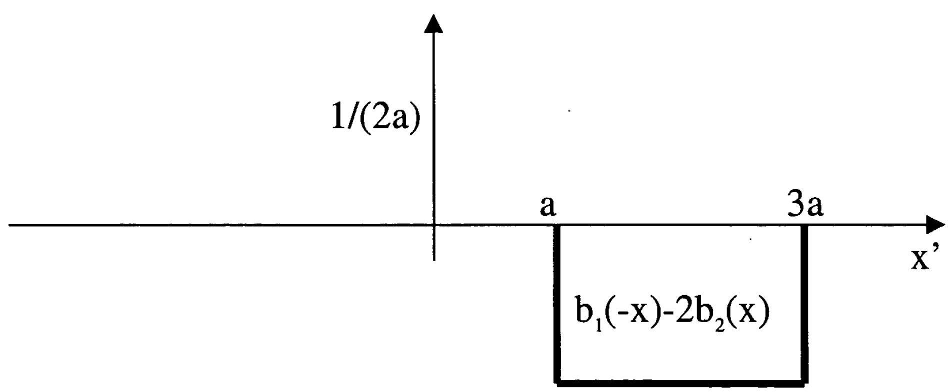


Fig 5B

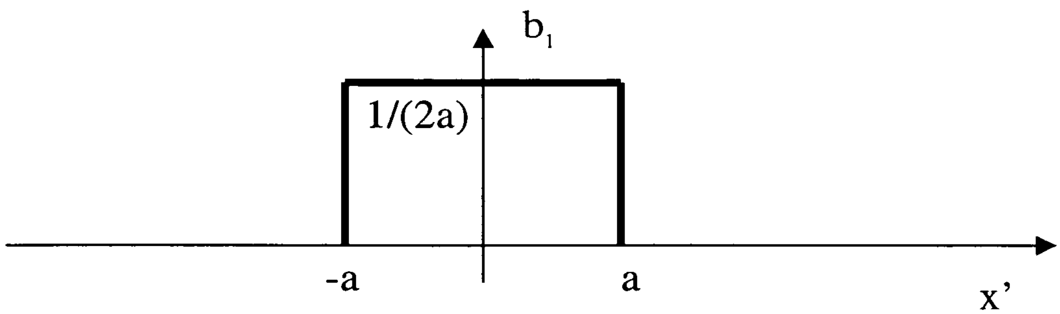


Fig 6A

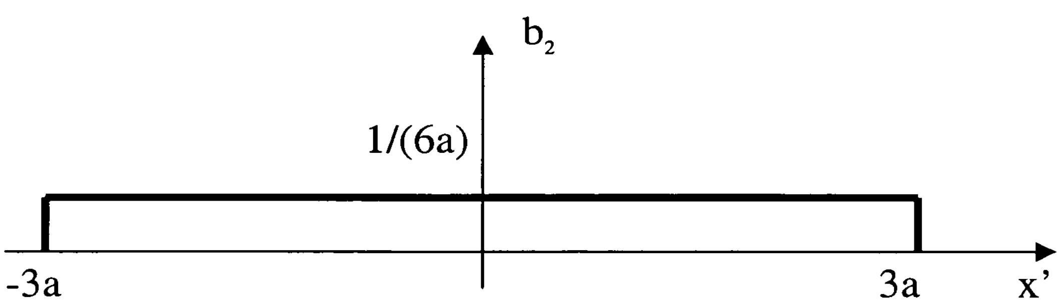


Fig 6B

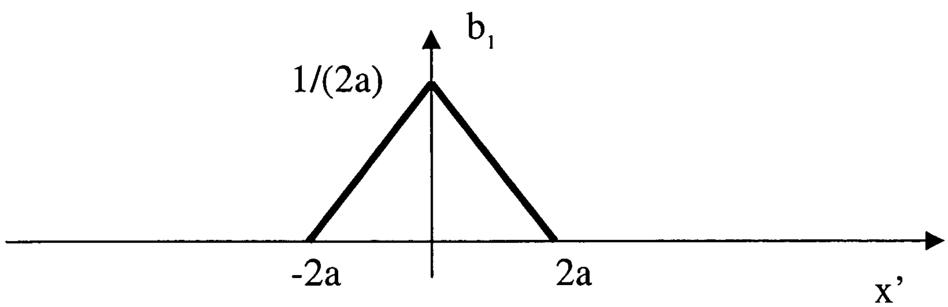


Fig 7A

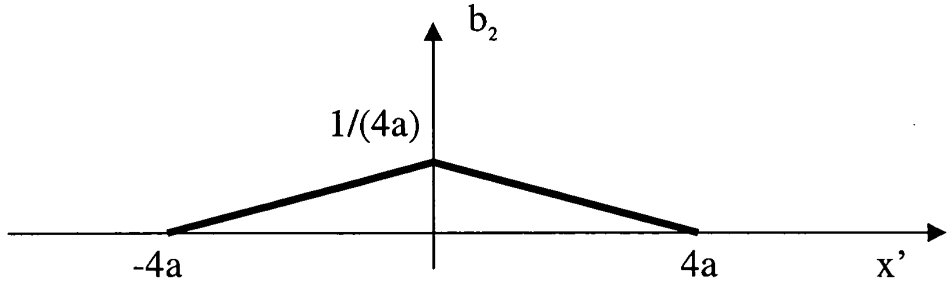


Fig 7B

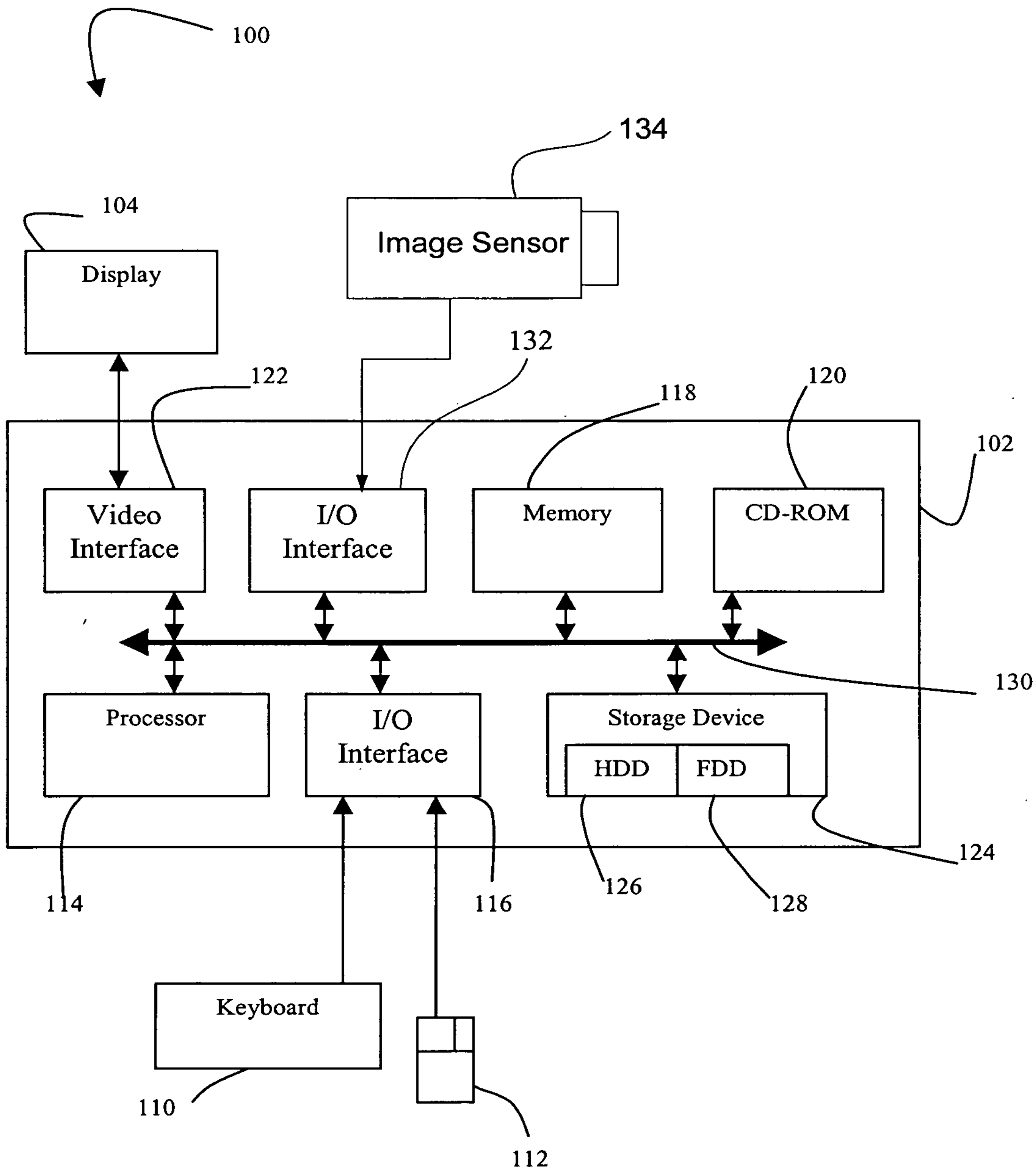

100
134
104
Image Sensor
Display
132
122
118
120
102
Video Interface
I/O Interface
Memory
CD-ROM
130
Processor
I/O Interface
Storage Device
HDD
FDD
124
114
116
126
128
Keyboard
110
112

Fig. 8

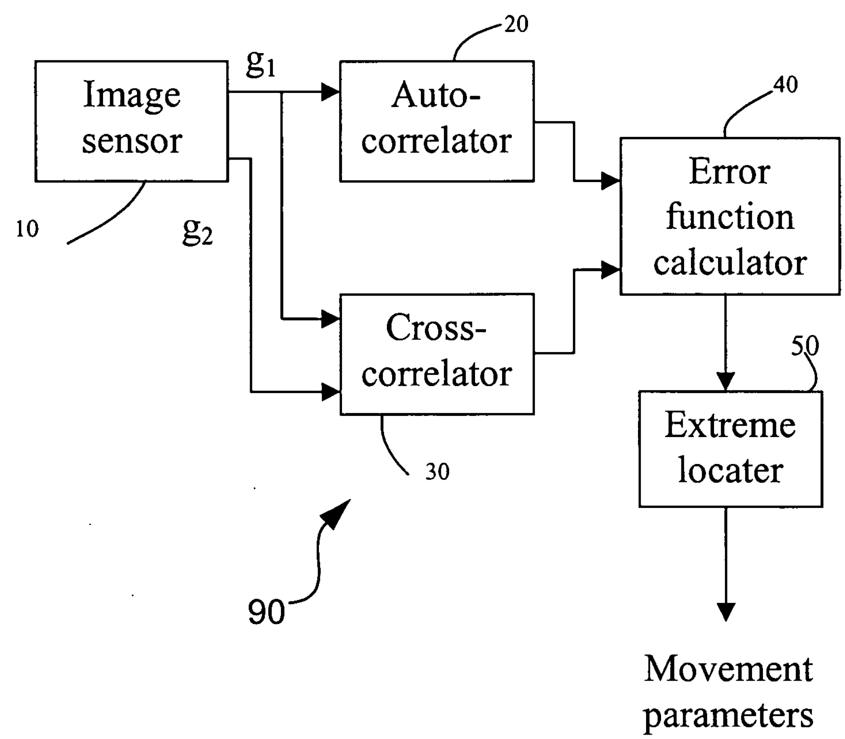

20
g_1
Image sensor
Auto-correlator
40
Error function calculator
10
g_2
Cross-correlator
50
Extreme locater
30
90
Movement parameters

Fig. 9

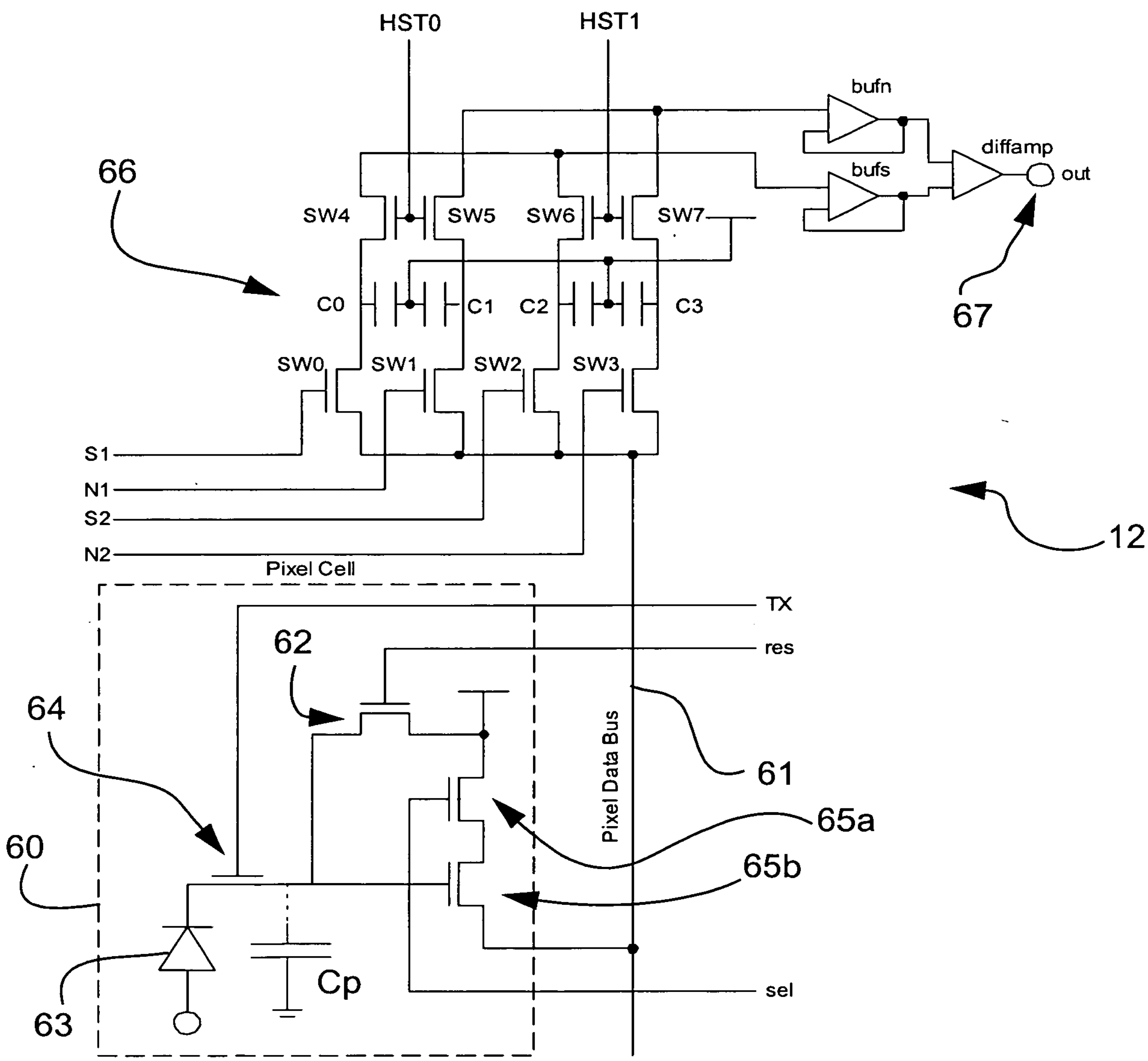
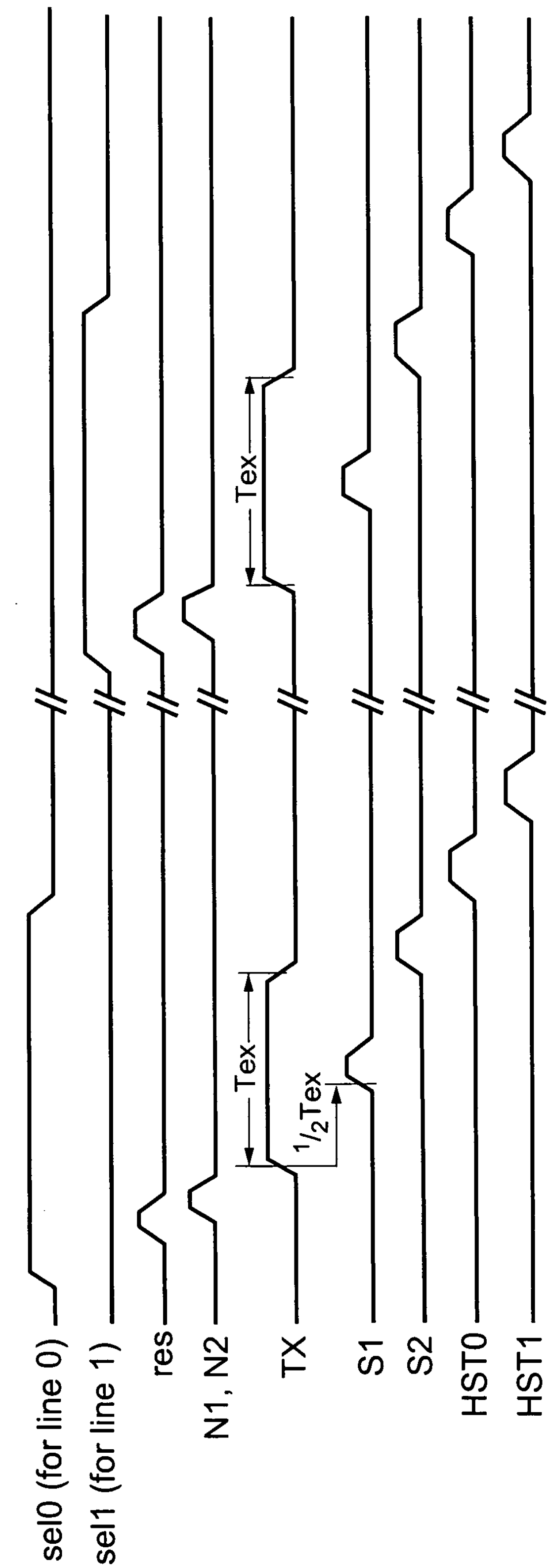

HST0
HST1
bufn
bufs
diffamp
out
66
SW4
SW5
SW6
SW7
C0
C1
C2
C3
67
SW0
SW1
SW2
SW3
S1
N1
S2
N2
12
Pixel Cell
TX
res
62
64
61
Pixel Data Bus
65a
60
65b
63
Cp
sel

Fig. 10



sel0 (for line 0)
sel1 (for line 1)
res
N1, N2
TX
S1
S2
HST0
HST1
Tex
Tex
$^1/_2$Tex

Fig. 11

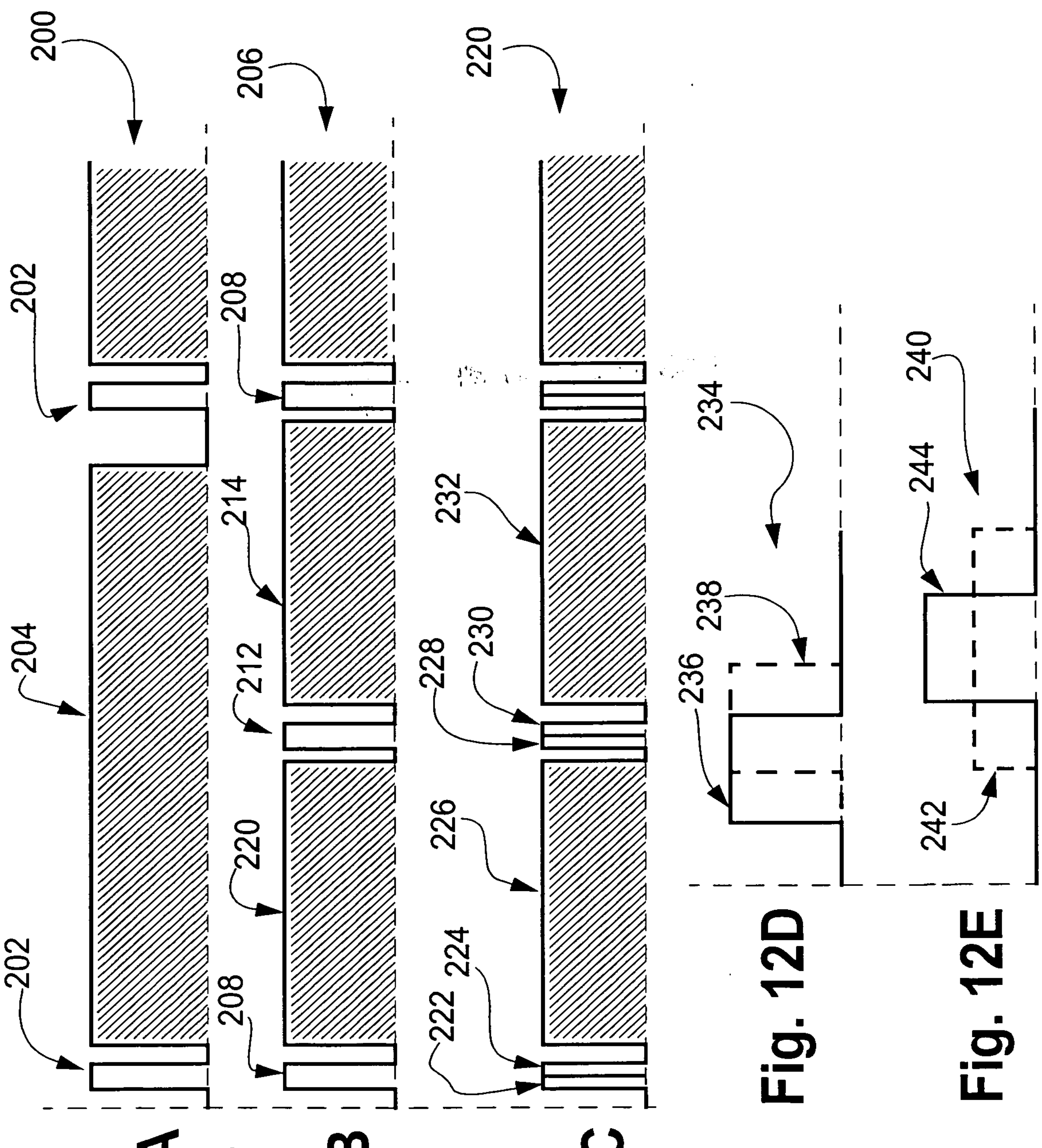


Fig. 12A (Prior Art)

Fig. 12B (Prior Art)

Fig. 12C

Fig. 12D

Fig. 12E